

Measuring syntactic variation in Dutch dialects

Marco René Spruit

Meertens Instituut, Amsterdam

<http://www.meertens.knaw.nl/medewerkers/marco.rene.spruit>

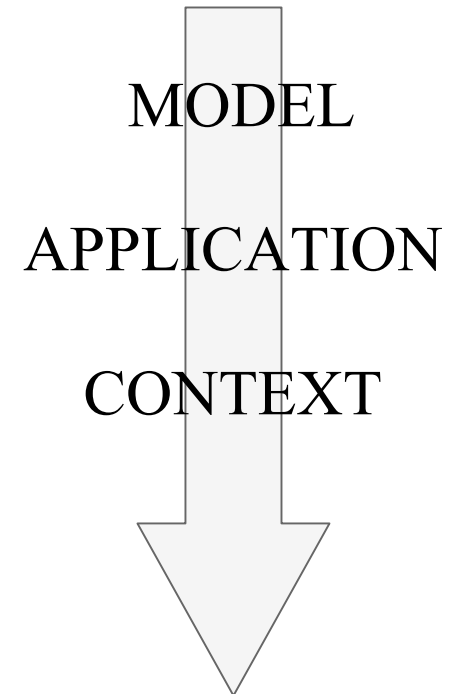
TIN-dag 2005

Context

- *Determinants of Dialectal Variation* project
 - Meertens: Hans Bennis, Sjef Barbiers
 - RuG: John Nerbonne, Wilbert Heeringa
- Correlation between geographical distance &
 - **Syntactic distance**
 - Phonological, lexical distance

Research questions

1. What is a syntactic measure?
2. **What are the syntactic distances between the Dutch dialects?**
3. How does the geographic distribution of syntactic phenomena in Dutch dialects relate to the distribution of dialect differences on other linguistic levels?



Dialectometry

- Measurement of dialect differences
 - i.e. linguistic differences whose distribution is determined primarily by geography
 - Initiator: Jean Séguy 1973
 - Canonizer: Hans Goebel 1982
- KEY: from measurement of individual linguistic variables to **aggregate** differences of varieties

Syntactic variation

- Syntactic Atlas of Dutch Dialects (SAND)
 - complementisers, subject pronouns, expletives, subject doubling, subject clitisation following yes/no, reflexive and reciprocal pronouns, fronting

Complementiser of comparative if-clause

't Lijkt wel **of** er
It looks AFFIX if there
iemand in de tuin staat.
someone in the garden stands.

*"It looks as **if** there is someone in the garden."*

{of, of dat, dat, as/of + embedded V2, at, as, et}

Weak reflexive pronoun as object of transitive verb

Toon wast **zich**.
Toon washes himself.

*"Toon washes **himself**."*

{hem, zijn eigen, zich, zichzelf,
hemzelf, zijn eigen zelve}

Data sample

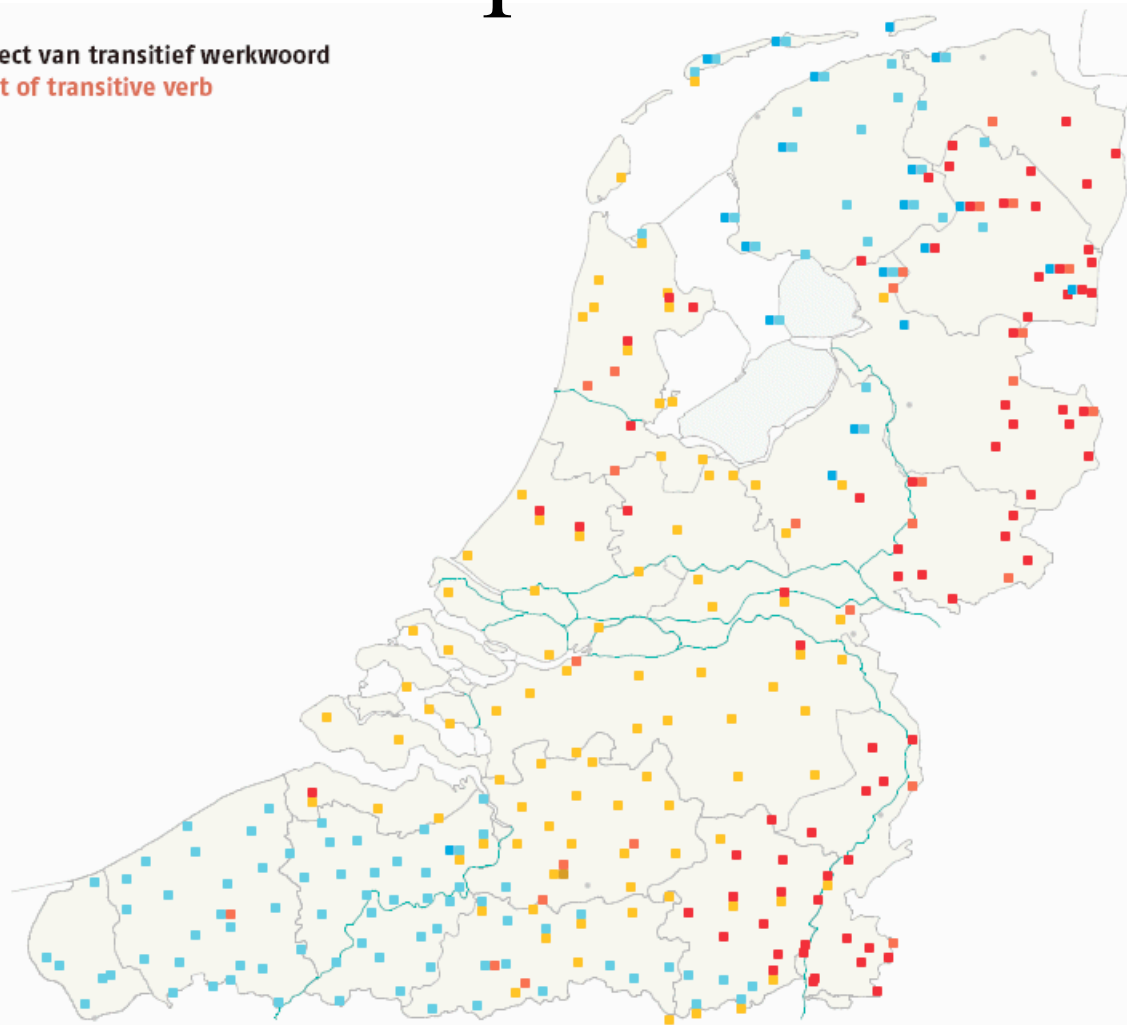
4.3.1.2 Zwak reflexief pronomen als object van transitief werkwoord

Weak reflexive pronoun as object of transitive verb

Toon wast zich.
Toon washes himself

'Toon washes himself.'

| | |
|--------------------|-----|
| ■ hem | 112 |
| ■ zijn eigen | 91 |
| ■ zich | 75 |
| ■ zichzelf | 26 |
| ■ hemzelf | 20 |
| ■ zijn eigen zelve | 1 |



Measurement

- Hamming distance
 - For each pair of dialects *A* and *B*:
 - For each **variant** of all syntactic phenomena:
 - If it **does occur** in dialect *A* but **not** in dialect *B*, or
 - If it **does not occur** in dialect *A*, but **does** in dialect *B*, then:
 - » Add 1 to the distance between dialect *A* and *B*

• Distance matrix:

| <i>DIALECT</i> | <i>A</i> | <i>B</i> | <i>C</i> | <i>D</i> | ... 267 |
|----------------|----------|----------|----------|----------|---------|
| <i>A</i> | | 1 | 3 | 6 | |
| <i>B</i> | 1 | | 2 | 5 | |
| <i>C</i> | 3 | 2 | | 3 | |
| <i>D</i> | 6 | 5 | 3 | | |
| ... | | | | | |

Visualisation

- Multidimensional scaling (MDS)
 - Procedure for producing a lower-dimensional data representation suitable for visualisation
 - Preserves some of the most prominent "distance" relationships
- MDS geographical dialect maps
 - Each dialect has its own unique RGB color
 - Color contrasts represent linguistic differences

Complementisers

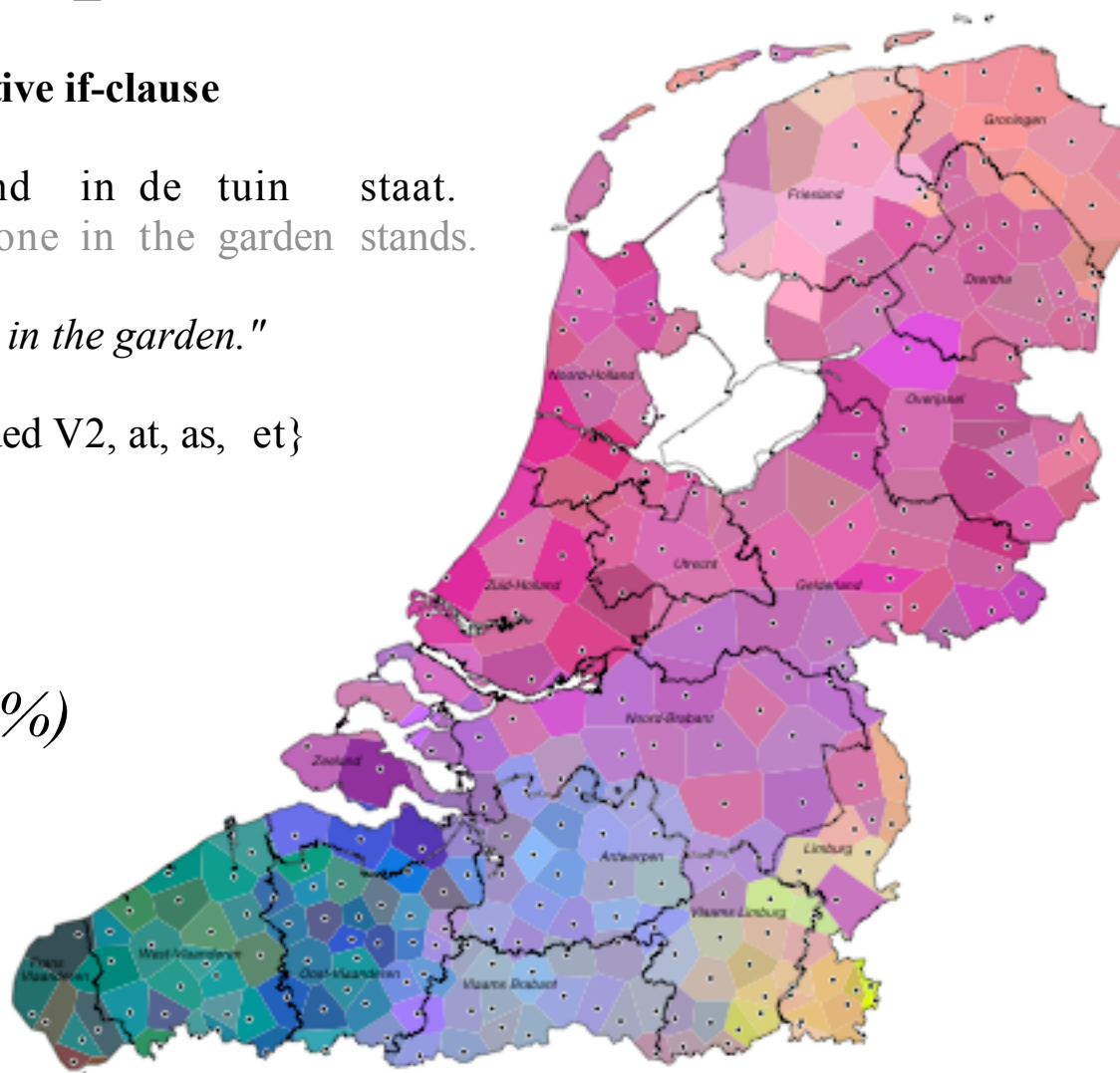
Complementiser of comparative if-clause

't Lijkt wel **of** er iemand in de tuin staat.
It looks AFFIX if there someone in the garden stands.

*"It looks as **if** there is someone in the garden."*

{of, of dat, dat, as/of + embedded V2, at, as, et}

- 101 variants (19.8%)
- $r = 0.94660937$



Subject pronouns

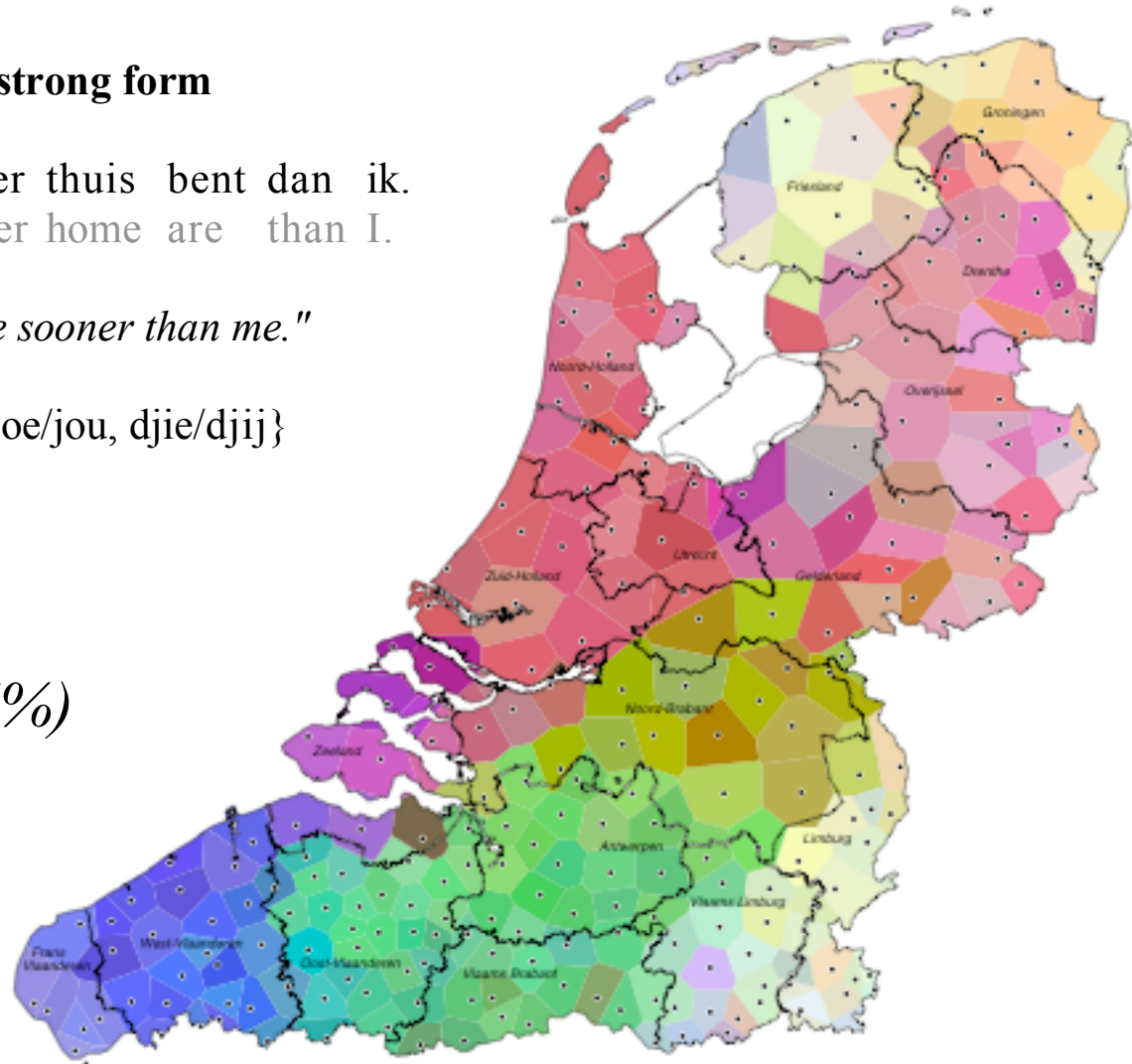
Subject pronouns 2 singular strong form

Ze gelooft dat **jij** eerder thuis bent dan ik.
She believes that you earlier home are than I.

*"She thinks that **you** 'll be home sooner than me."*

{gij/gie, jij/jie, du, ie, dich, jo/joe/jou, djie/djij}

- 172 variants (33.7%)
- $r = 0.88065714$



Expletives

Expletive subject following complement in existential clause

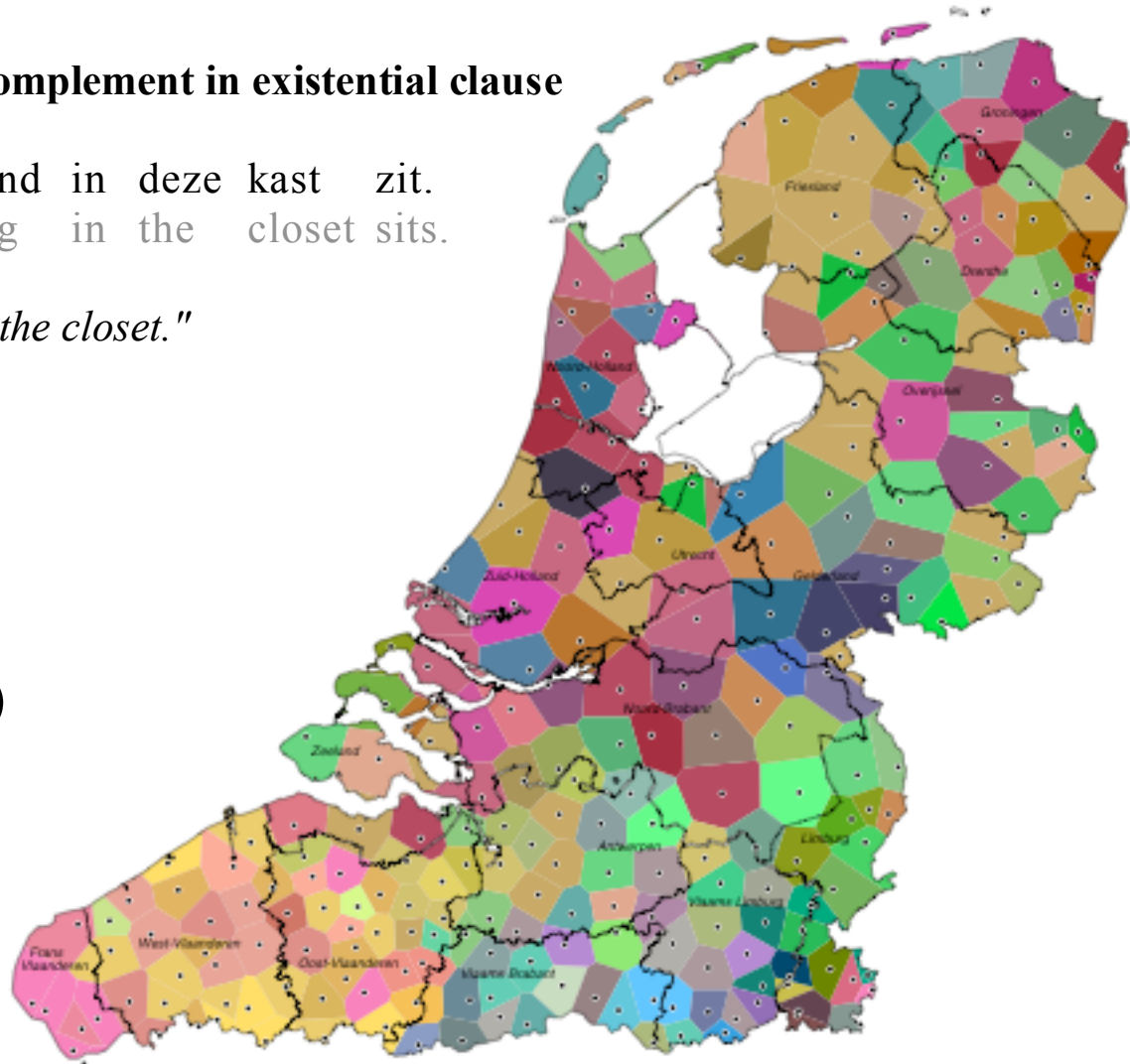
't Is net of er een hond in deze kast zit.
It is just if there a dog in the closet sits.

*"It looks as if **there** is a dog in the closet."*

{er/d'r/t'r, daar, Ø}

- 13 variants (2.5%)
- $r = 0.8739387$

mrs2005tin



Subject doubling

Subject doubling 2 singular

As-ge **gij** gezond leeft,
If-you_{WEAK}-you_{STRONG} healthily live

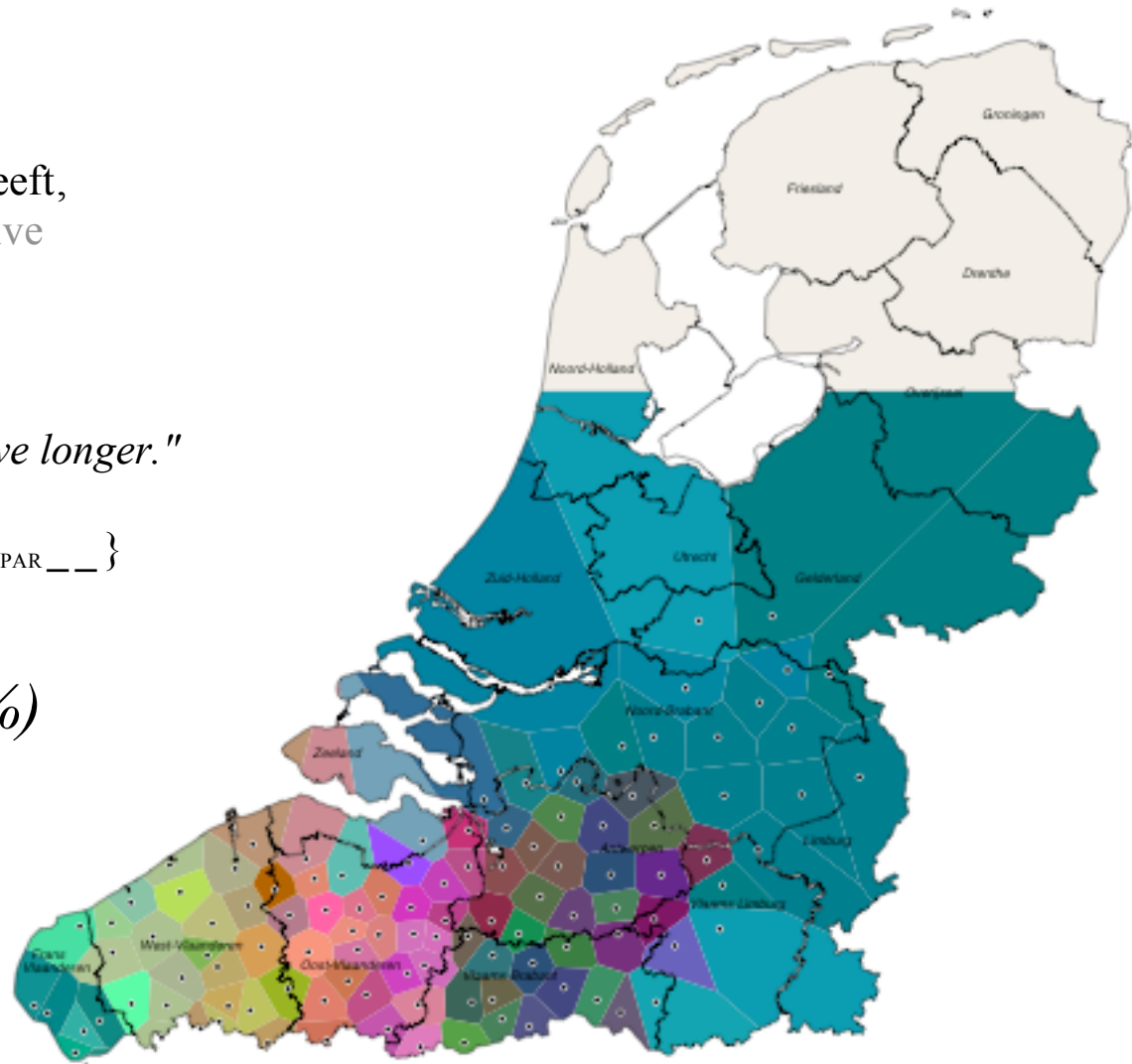
leef-**de gij** langer.
live-you_{WEAK}-you_{STRONG} longer.

*"If **you** live healthily **you** will live longer."*

{V_{FIN} __, __ V_{FIN} __, C __, C_{COMPAR} __}

- 54 variants (10.6%)
- $r = 0.95438211$

mrs2005tin



Subject clitisation following yes/no

Clitics and agreement following ja/nee 'yes/no' 3 plural

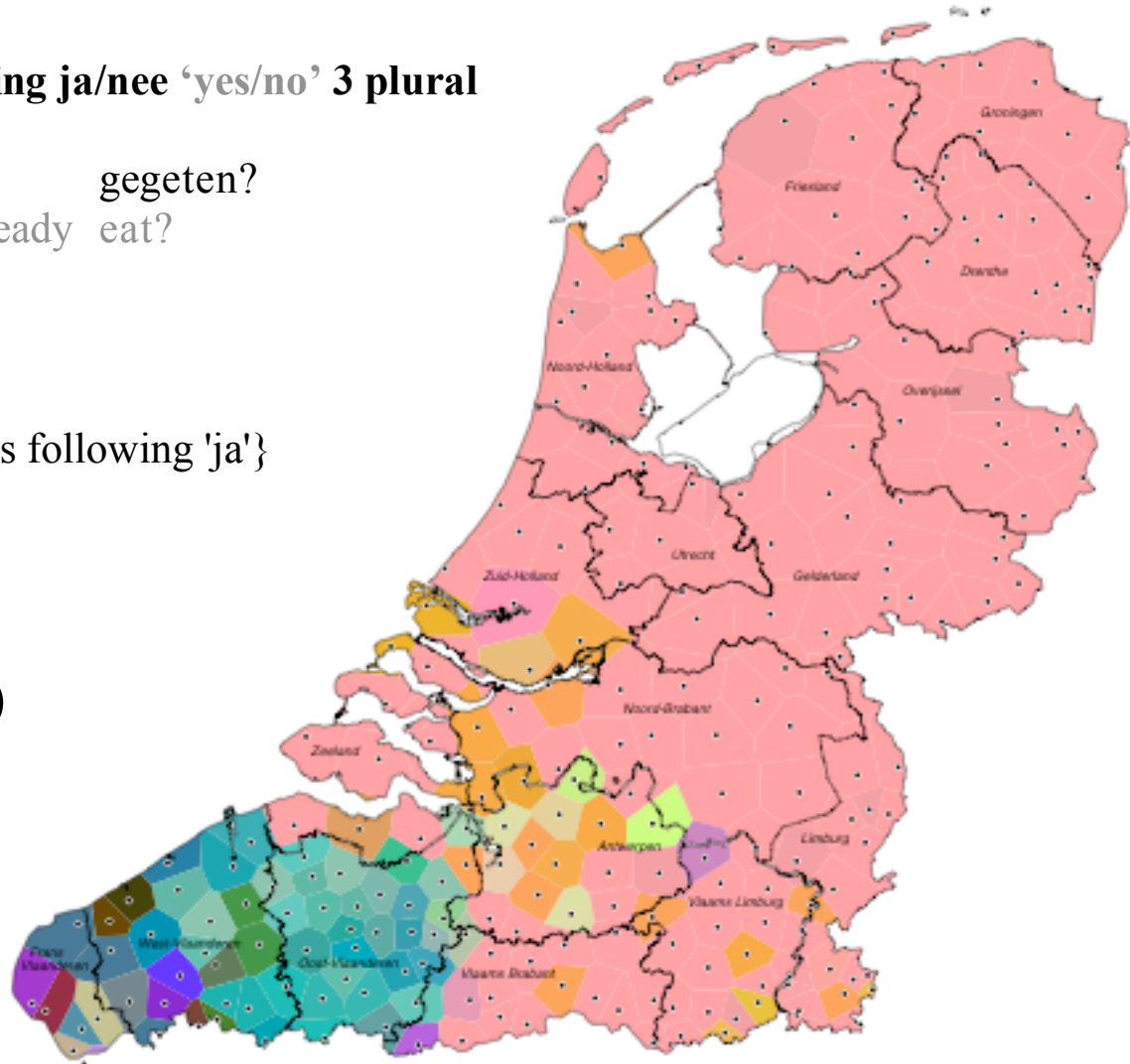
Vraag: Hebben ze al geeten?
Question: Did they already eat?

Antwoord: *Jaa-se*

Answer: Yes-3PLUR

{jaa-s(e), jaa-ns, jaa-t, no clitics following 'ja'}

- 30 variants (5.9%)
- $r = 0.99025193$



Reflexive & reciprocal pronouns

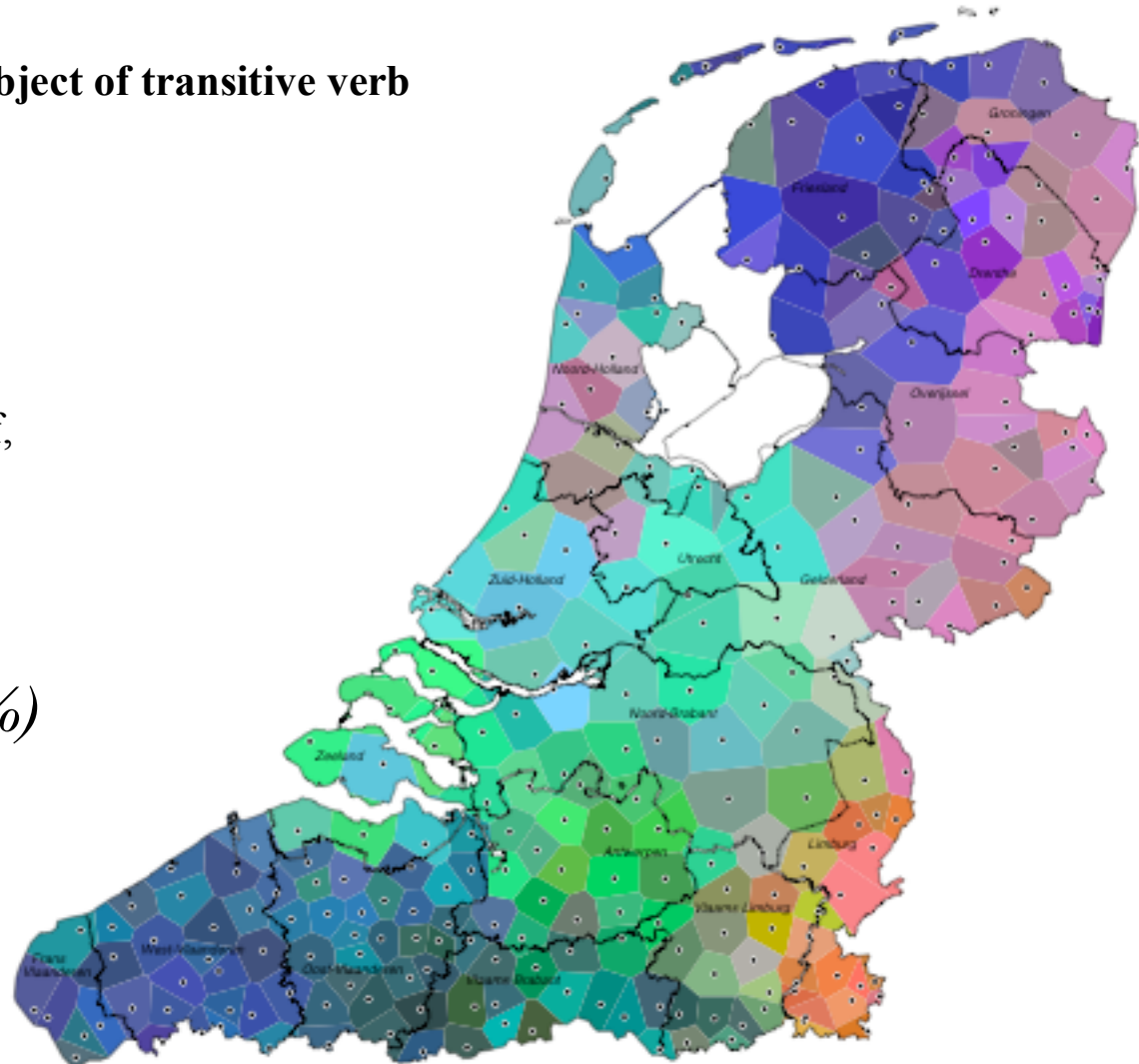
Weak reflexive pronoun as object of transitive verb

Toon wast **zich**.
Toon washes himself.

*"Toon washes **himself**."*

{hem, zijn eigen, zich, zichzelf,
hemzelf, zijn eigen zelve}

- 78 variants (15.3%)
- $r = 0.93453301$



Fronting

Short object relative

Dat is de man **die** ze geroepen hebben.
That is the man who they called have.

*"That is the man **who** they have called."*

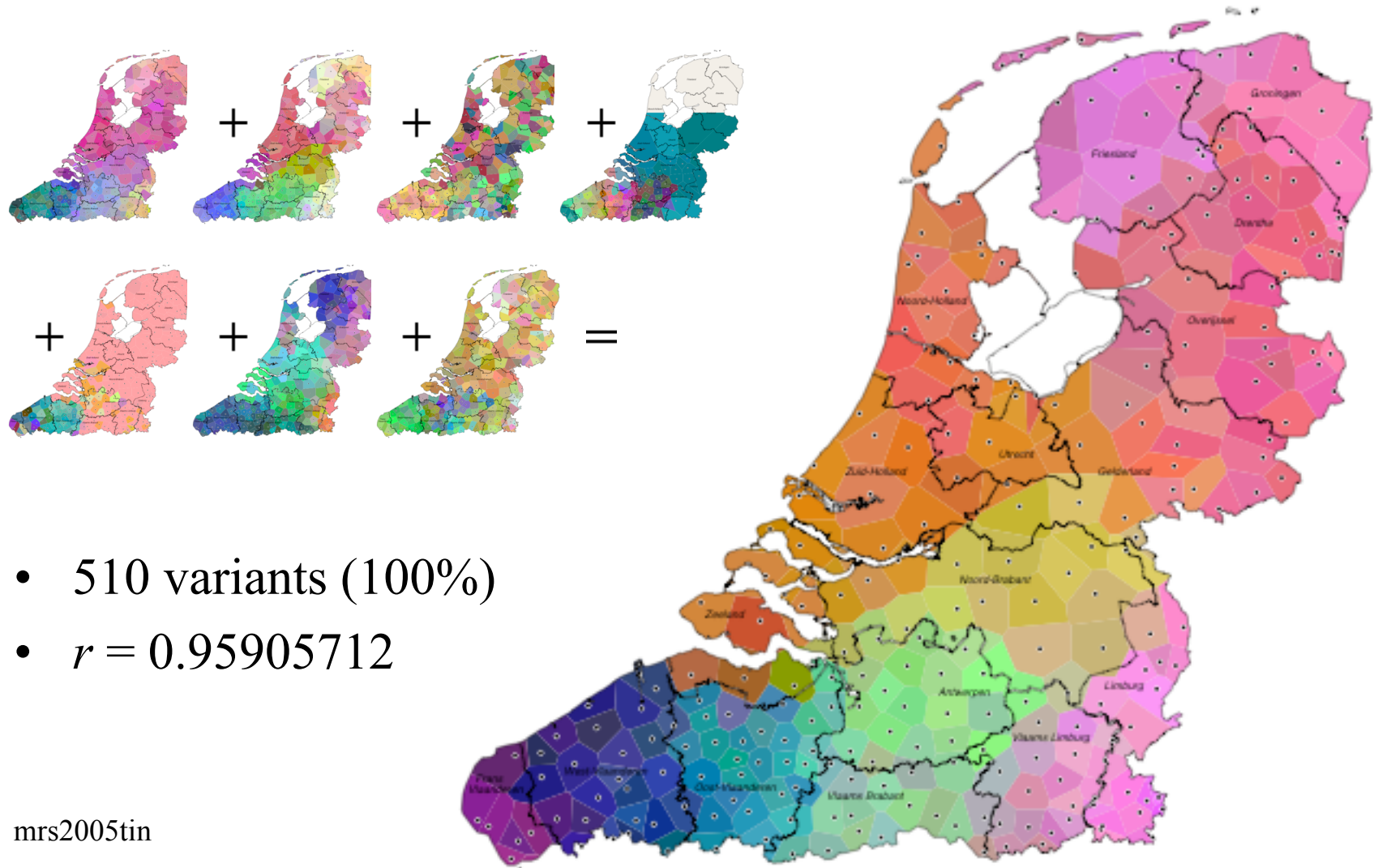
{die, dat, wie, der, den/dem, as}

- 62 variants (12.2%)
- $r = 0.77975377$

mrs2005tin

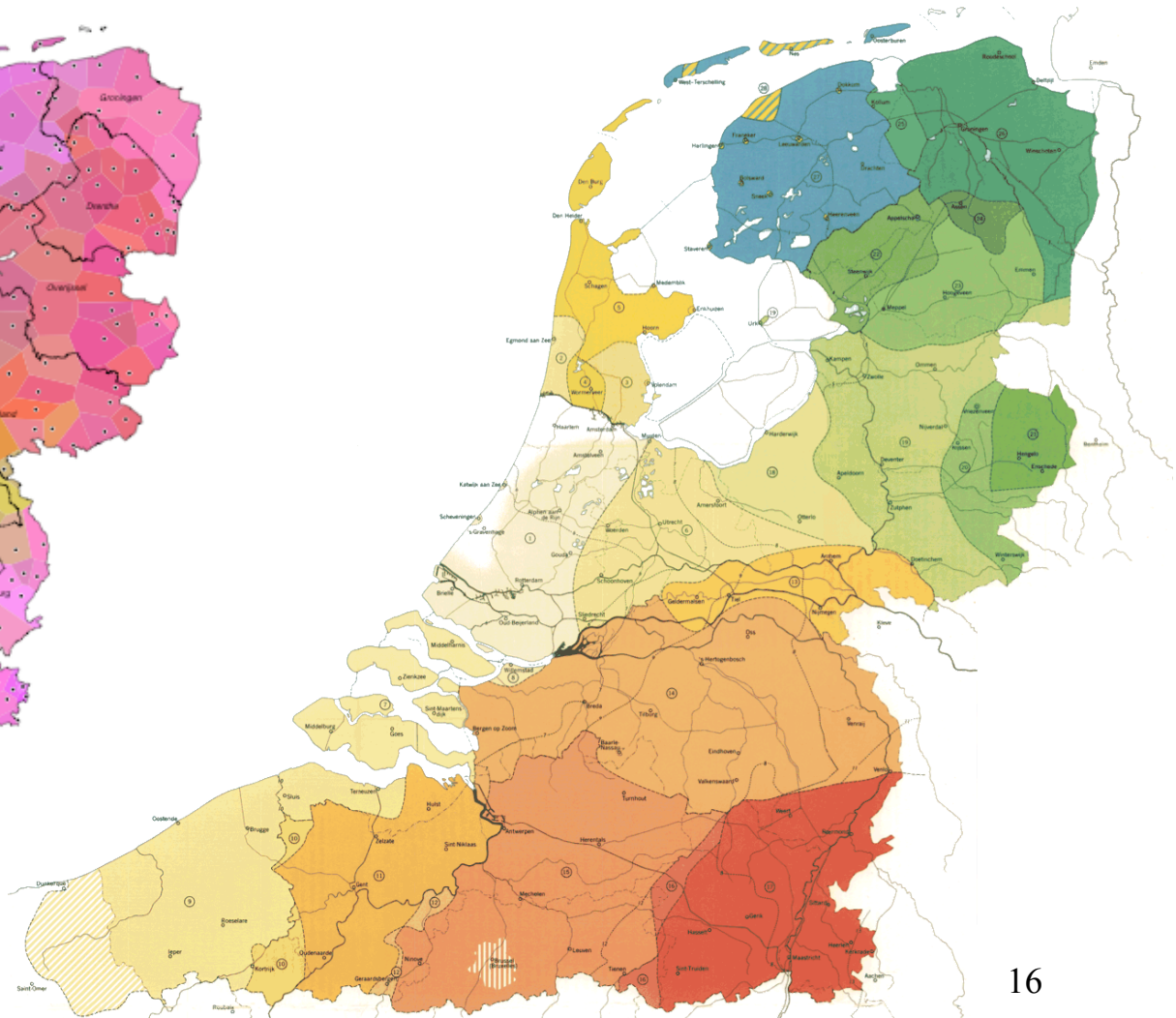


SAND map

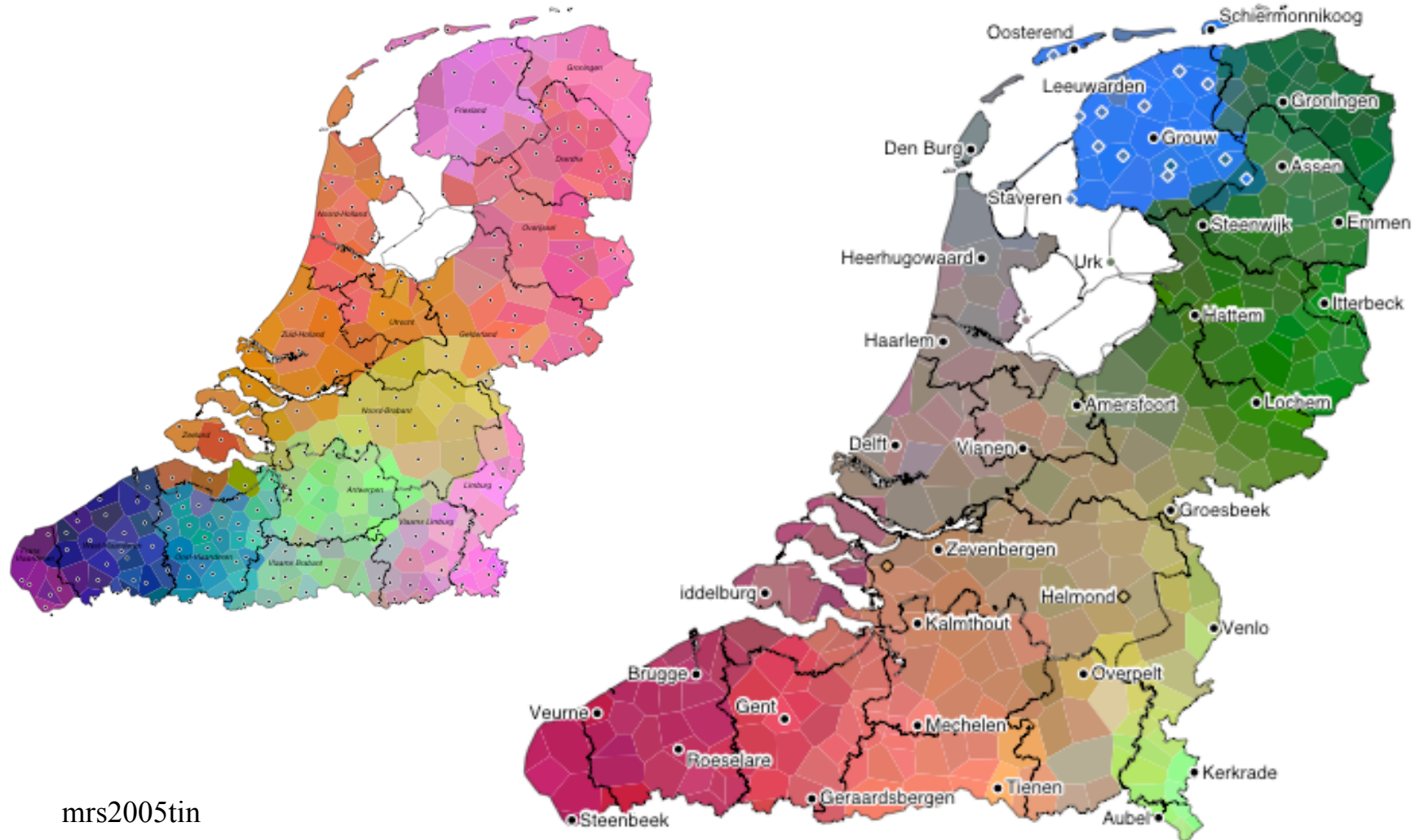


mrs2005tin

Daan & Blok dialect map 1969



Heeringa dialect map 2004



mrs2005tin

Future research

- More advanced measurements
 - Incorporation of variant frequency (G.I.W.)
 - Accounting for variant probability (alternative variants for a feature)
- Exploring dependencies between syntactic phenomena